Underwater 3D Structures As Semantic Landmarks in SONAR Mapping

Underwater 3D Structures as Semantic Landmarks in SONAR Mapping

Thomas Guerneve[1], Kartic Subr[2] and Yvan Petillot[3]

***Abstract*— SONAR mapping of underwater environments leads to dense point-clouds. These maps have large memory footprints, are inherently noisy and consist of raw data with no semantic information. This paper presents an approach to underwater *semantic mapping* where known man-made structures that appear in multibeam SONAR data are automatically recognised. The input to the algorithm consists of SONAR images acquired by an Autonomous Underwater Vehicle (AUV) and a catalogue of 'guessed' 3D CAD models of structures that may potentially be found in the data. The output of our algorithm is online 3D mapping, with navigation correction. In addition, for any objects in the input catalogue, the dense point clouds of those objects are replaced with the corresponding CAD model with correct pose. Our method operates with a catalogue of coarse CAD models and proves to be suitable for online semantic mapping of a partially man-made underwater environment such as a typical oil field. The semantic world model can finally be generated at the desired resolution making it useful for both offline and online usual processing such as mission planning, data analysis, manipulation or vehicle relocalisation.**

Our algorithm proceeds in two phases. First we recognise objects using an efficient, rotation-invariant 2D descriptor combined with a histogram-based method. Then, we determine pose using a 6 degree-of-freedom registration of the 3D object to the local scene using a fast 2D correlation which is refined with an iterative closest point (ICP) -based method. After structures have been located and identified, we build a semantic representation of the world resulting in a lightweight yet accurate world model. We demonstrate the applicability of our method on field data acquired by an AUV in Loch Eil, Scotland.

I. INTRODUCTION

Inspection and mapping of underwater environments is challenging. Navigation through these environments relies heavily on the quality of mapping and on identification of subsea features since Global Positioning Systems (GPS) cannot be used underwater. Due to the scattering of light underwater, developments in computer vision techniques for detecting and recognising objects from optical images are not applicable in most practical scenarios. Acoustic imaging using SONAR is the most popular approach. While there has been much activity in developing low-level tools for processing SONAR images ([16], [24]), there is an imminent need for performing high-level tasks such as object detection and recognition. These tools will allow information to be gleaned from dense arrays of SONAR images so that maps may be represented semantically in terms of known objects rather than as pixels.

[1]SeeByte Ltd and Ocean Lab Sytem, Heriot Watt University, Scotland - tsg1@hw.ac.uk

[2]University of Edinburgh, Scotland - K.Subr@ed.ac.uk

[3]Heriot Watt University, Scotland - Y.R.Petillot@hw.ac.uk

In preparation for SONAR mapping using AUVs, very little is known about the environment *a-priori*. For example, the existence of the types of man-made structures may be guessed and their positions and orientations may be known approximately (± 7 metres and $\pm 30^\circ$ respectively). Operating around these man-made structures requires careful planning and stand-off distances for safety, limiting the resolution of the observations. The resulting uncertainty on the prior map increases the complexity of operations since compensatory mechanisms for error are costly and due to the need for robust relocalisation methods.

While prior 3D mapping of the site provides a solution to the navigation problem by creating occupancy grids, enabling online obstacle avoidance and path planning, complex operations often require an analysis of the 3D map by a human operator to identify the elements of interest and the potential hazards [9]. This intervention considerably increases the operation time and stress on the human operator. In this context, the ability to perform an online semantic mapping both improves time-efficiency and safety by providing a representation adapted to the complexity of scene and focused on the elements of interest to the operators.

Simple and stereo optical mapping require pristine visibility conditions and are restricted to the observation of textured objects. Laser-based systems provide accurate range measurements but typically feature low sensor footprints resulting in either very sparse sampling or increased scanning time. In both cases, these solutions are sensitive to water turbidity and restricted to short range observations. Unlike optical sensors, acoustic sensors enable weather-independent long-range sensing at the cost of lower spatial resolutions. While the accuracy of 3D from SONAR is typically limited by the vertical aperture of the sensor, previous work ([7], [1]) demonstrated the ability to obtain reconstructions at centimetre-level accuracy using standard imaging SONARs. Inspired by this, we exploit such reconstructions to enable visual object recognition for semantic mapping.

We present an online, semantic modelling pipeline of worlds based on imaging SONAR data and coarse prior information on potential objects of interests. Our simple method will enable an online simultaneous mapping and navigation correction, akin to SLAM [22]. The prior information is supplied as a catalogue of 3D CAD models of objects that are likely to be found. We introduce an efficient method for object recognition based on a rotation-invariant descriptors. Then, we develop a two-step robust model registration technique that provides centimetre-level localisation. Finally, we validate our method using experimental results on field data acquired by an AUV in Loch Eil, Scotland.

II. LITERATURE REVIEW

Semantic representations of environments offer multiple advantages [13]: Lightweight representation enables online reporting [10]; identification of objects of interest enable autonomous manipulation such as target recovery [14]. We provide a brief overview of underwater mapping and semantic mapping from SONAR data.

A. Underwater SLAM

Underwater SLAM from SONAR sensors has been widely studied, both in man-made environments [22] where the walls of a harbour can be used as features or based on the seabed geometry such as in [18] where a probabilistic ICP-based registration yields results comparable to USBL localisation. The registration process can be performed either on the whole data such as with ICP-based methods [25] or on selected features ([6], [2]). When using features, the high variability of acoustic returns based on the environment (type of seabed, multi-path) and the position of observation often impose manual input [8]. On the other hand, direct registration techniques offer the interest of taking into account the whole information at the cost of higher sensitivity to noise and increased resource usage.

In light of this, our feature-less registration technique propose a robust approach by relying on a percentile of the data and is resource-efficient by adopting a lightweight 2D elevation map representation.

B. Semantic mapping from SONAR

Although widely investigated in the air-domain ([19], [21]), semantic mapping in the underwater domain has been mainly restricted to the study of seabed [5] and natural habitat [12]. Due to the difficulty of acquiring accurate underwater 3D representation, most semantic mapping studies focus on backscatter classification [17]. However recent research work [7] showed that multibeam SONAR provided centimetre level reconstruction accuracy, allowing geometry-based semantic interpretation. Following the idea of adopting a plane-based representation [4], the authors of [20] presented a two-level semantic annotation of the environment. Planes are first fitted to operate a rough classification of the areas into general categories (seabed, wall) while a second step focuses on the type of object by analysing the local normals distributions.

We propose to take advantage of the spatial accuracy of 3D reconstructions from SONAR to operate a geometrical object recognition from a set of CAD models describing the structures of interest. Our approach proposes to recover the 6D position of the objects of interest in an operation time, enabling online inspection planning and relocalisation.

III. METHOD

Our world modelling pipeline is composed of four main steps. First the field area selected for inspection is reconstructed in 3D using multibeam SONAR observations and vehicle navigation data. Once the full field representation available, a model-based structure recognition is performed using a circular histogram-based description scheme. In order to do this, a set of CAD models representing the structures of interest is used as prior information. Following the recognition of all the structures of interest, a fine registration of their associated model is performed in two steps. First each reconstructed structure is isolated and its associated CAD model is matched in 4D (*North, East, Depth, Yaw*). A standard ICP-based method is then used to refine the initial matching and allows for correction of possible offsets in the remaining two dimensions (*Roll, Pitch*). Once all the structures identified and registered, the reconstructed scene is segmented into two classes respectively representing the structures of interest and the seabed. A lightweight world model representation is then generated using the set of registered CAD models and a surface representation of the seabed.

A. Field reconstruction

As depicted in fig. 1, a typical lawnmower pattern is first followed by the AUV, featuring multiple swathes with significant overlap. A 3D reconstruction of the full field is

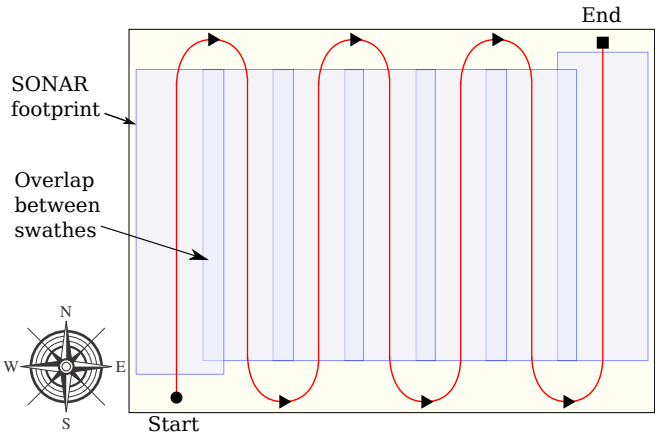


Fig. 1. Lawnmower inspection pattern followed by the vehicle (red) to inspect the field area (yellow). The multibeam SONAR footprint is represented in blue, showing overlapping regions between each consecutive swathes used to reduce the navigation drift occurring after each turn.

then obtained using a 2D imaging SONAR and navigation data. The 3D reconstruction technique is based on a carving approach [7] and assumes good local navigation. While the navigation information is commonly assumed to be accurate enough over a few metres (typically less than a percent of the distance travelled), significant navigation drift can be observed when following sharp turns such as U-turns or long swathes. In this situation, a registration between multiple observations of the scene provides a way to compensate for navigation errors. Therefore and in order to obtain a consistent 3D representation of the full field, we extended the work presented in [7] with a 3D registration step.

The field reconstruction is thus divided in multiples swath patches registered to each-other using the overlapping depth measurements. Since compass sensor readings are typically very reliable (a fraction of degree accuracy), this bundle adjustment process is operated in 3D only (*North,East,Depth*). In this situation, a suitable representation is given by the use of 2D elevation maps generated at a given *North/East* resolution.

Each new elevation map is then registered to the previous one by minimizing their unsigned median *Depth* euclidean distance, normalized over the number of overlapping pixels. Importantly and in order to account for potential elevation offsets due to SONAR miscalibration and pressure sensor inaccuracy, the depth distributions of the overlapping submaps are previously aligned by matching their signed median values. The final 3D map is then given by combining the co-registered 3D swathes.

B. Structure recognition

1) Description: Once the whole field is mapped in 3D, a set of CAD models is used to describe the structures expected to be found on the field. The structures are assumed to be laid on the seabed with arbitrary *Yaw* orientations. In order to reduce dimensionality, an elevation map representation is adopted to describe both the field and the model (see fig. 2-a). So as to decouple the 3D model information from the *Yaw* orientation, a rotation-invariant description scheme is adopted by computing elevation values histograms on N complementary circular regions, as depicted on fig. 2-b-g. An histogram of the elevation values of the CAD model is

Fig. 2. a) Elevation map of a CAD model. b-g) A rotation invariant description of the CAD model is obtained by computing histograms on multiple (N = 6 in this example) complementary circular areas.

then generated on each circular region, resulting in an $N \times P$ descriptor with P being a number of bins of the histograms, representing the depth resolution.

2) Recognition: As illustrated in fig. 3, the structure recognition algorithm is based on a histogram comparison on a set of of complementary circular areas. Due to the necessity of covering the full field, a divide-and-conquer approach is adopted. An elevation map representation of the reconstructed field is thus divided in multiple 2D circular patches, extracted at a given sampling period Δ_{NE}. The radius of the patches r_{patch} is then defined based on the sampling period and the size of the considered model r_{model} by $r_{patch} = r_{model} + \frac{\sqrt{2}}{2}.\Delta_{NE}$, with r_{model} being the largest distance between the centre of the model and one of its points. Following this sampling scheme, at least one patch is guaranteed to contain the full reconstructed structure. The recognition step aims at finding this patch by selecting the patch that resembles the most to the model. Following the same method as described in section III-B.1, a circular histogram-based description is employed to describe each patch. When comparing them to the CAD model elevation map, the depth of this latter is first defined by aligning their elevation distributions using the same technique described in section III-A. Once the depth of the CAD model set, the Earth Mover's Distance metric [23] is used to characterize their similarity on each circular region. The distances computed on each circular patches are concatenated in a vector. The

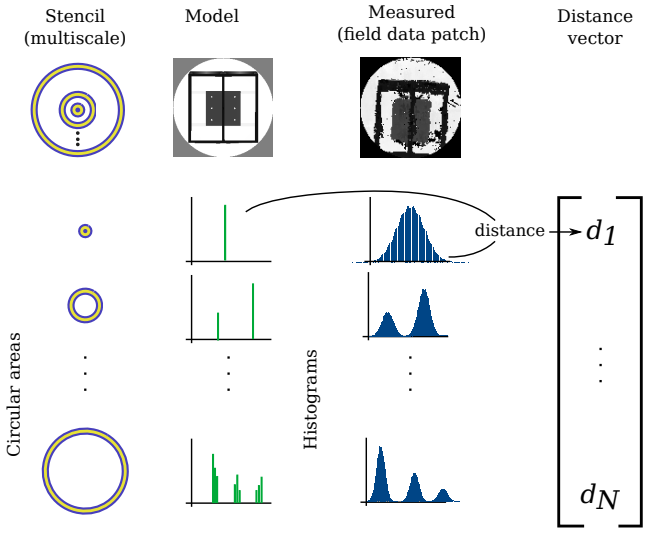


Fig. 3. Structure recognition technique. Each field subdivision is compared to the model using a histogram comparison for which we use Earth Mover's Distance (EMD). The distances are computed on a set of multiscale circular areas to ensure rotation invariance. The norm of the distance vector of EMD distances $[d_1 \dots d_N]$ is used as a similarity score.

euclidean norm of this vector is chosen as the similarity score. The candidate patch with the lowest score is finally selected as being the best representation of the model in the scene.

C. Fine model matching

Once the structure located, the CAD model is finely registered in the patch selected during the previous step with a two-step registration procedure.

Due to the small size of the patch, an exhaustive 4D matching is performed, considering a discrete set of translations along the North and East axis and structure orientations in $[0°, 180°]$ or $[0°, 360°]$ angular intervals, depending on the symmetry of the object. The depth of the model is set following the same approach as described in section III-A, based on the signed median elevation. The optimal registration is then chosen as the minimum euclidean distance between the transformed model and the patch.

In order to account for possible offsets in the remaining two dimensions (non-flat seabed), a robust ICP-based method is applied [3], [15] by discarding the furthest points (outliers) for registration.

D. World model generation

Once all the models registered to the scene, a simplified representation of the scene is generated by replacing the reconstructed structures by their registered CAD model representations. The remaining points can then be considered as representing only the seabed and a surface is then reconstructed by interpolation [11] at the desired level of accuracy.

In addition to this 3D representation, a graph-based representation is generated, retaining the position of the structures of interest as well as their potential connectivity information when detecting the presence of a pipeline between two structures.

E. Algorithm optimizations

When large areas are inspected, potentially featuring a large number of structures, the computational cost can be reduced by optimizing the modelling process. In order to do this, the structures are detected in the scene sequentially and in decreasing size order, allowing the removal of large areas of the field after their detection. The area to process therefore decreases after each new detection.

Since our recognition approach is based on the minimization of a similarity score, estimating a lower bound of this score enables to discard wrong solutions at an earlier stage. Our similarity score being an euclidean distance of an N-dimensional vector, the temporary distance given by computing the distance on k elements $(1 \leq k \leq N)$ naturally provides a lower bound to the similarity distance. We therefore estimate the similarity of the patches at each step by computing partial distances with increasing k numbers, aborting the comparison earlier when the distance appears larger than the current optimum.

IV. EXPERIMENTAL RESULTS

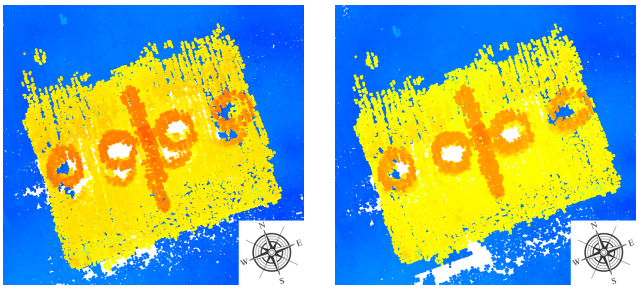
We present in this section field experimental results based on a dataset acquired in Loch Eil in Scotland. The data was gathered by the Subsea 7 AIV (Autonomous Inspection Vehicle) prototype during validation trials. Prior to the experiment, three oil-field type structures were laid on the seabed with various orientations. The field area was approximately 50×34 metres. The vehicle was equipped with a BlueView MB2250 SONAR mounted in a downward configuration featuring a pencil-beam vertical aperture of 1° and an horizontal aperture of around 80°. The along-track sampling period was 4cm with an average range to the seabed of 5m. The navigation data was provided by a module integrating the readings of three sensors : bottom lock velocities were provided by a DVL (Doppler Velocity Log), depth measurements were given by a pressure sensor while the estimation of the vehicle orientation relied on a compass and a gyroscope.

A. Field reconstruction

The field in Loch Eil was inspected following a lawnmower pattern with swathes along the *East-West* axis. The swathes elevation maps were generated at 1cm resolution in *North/East*. Since the along-track sampling period is only 4cm on average, some gaps are apparent on the map. The bundle adjustment was therefore operated at 1cm resolution, discarding undefined points. To illustrate the benefits of the registration between swathes, fig. 4 shows the 3D reconstruction of the structures before (a) and after registration (b) leading to a shift of 24cm in *North* and 2cm in *East*. While reconstructing the whole field, the maximum registration shift was 38cm in *North* and 18cm in *East*.

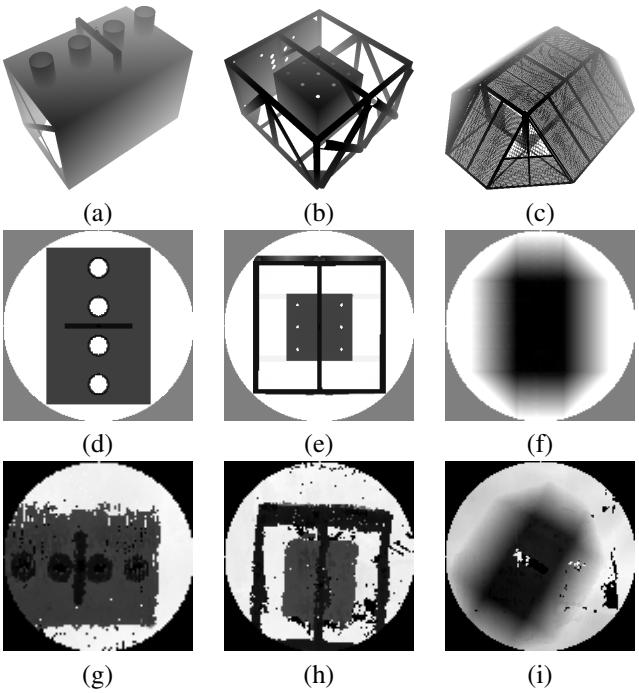
B. Structure recognition

Once the full field area reconstructed, 3 rough CAD models (see fig. 5-a-c) of the structures present in the field are used as prior information of the structure recognition step. The



(a) (b)

Fig. 4. Illustration of color-coded (based on the altitude) 3D point clouds. a) 3D reconstruction of a structure without registration. b) Registering the two swathes containing the structure enables to recover a consistent 3D reconstruction.



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig. 5. Model-based structure recognition. a-b-c) The CAD models used as prior information for the object recognition step are converted to elevation maps (d-e-f) to perform a structure recognition step on the full field, leading to the selection of the three most similar patches representing the structures of interest (g-h-i).

sampling period Δ_{NE} used for the patch generation was 8cm. As can be seen in fig. 5-g-i, the recognition algorithm selected the correct patches based on their similarity to their elevation maps (fig. 5-d-f).

C. Fine model registration

As shown in fig. 6-a-c, the 4D matching based on the elevation maps provide a registration on a fixed grid (8cm period in *North* and *East* and *Yaw* angles steps of 5°) exhibiting small angular and translation offsets. The ICP-based registration step then corrects for the remaining gaps as can be observed in fig. 6-d-f. Table I gathers distance metrics between the registered point cloud and their CAD models, exhibiting an average 40% improvement in registration when using ICP with a final average median error under 3cm.

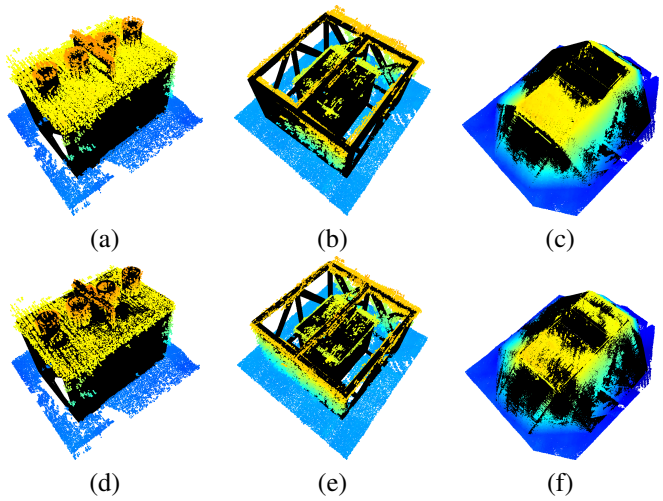


Fig. 6. Two-step model registration. a-b-c) The 4D registration based on the elevation map representation provides a first rough registration of the model on a fixed grid. d-e-f) The second registration step, based on ICP, provides a full 6D registration.

	Structure 1	Structure 2	Structure 3
Step 1	0.083 (0.086)	0.052 (0.075)	0.027 (0.041)
Step 2	0.024 (0.037)	0.041 (0.054)	0.018 (0.026)

TABLE I

UNSIGNED MEDIAN DISTANCES (AND MEAN DISTANCES) IN METRES BETWEEN THE RECONSTRUCTED STRUCTURES AND THEIR CAD MODELS AFTER EACH REGISTRATION STEP.

D. World model generation

Following the registration of all the structures, a simplified world model is generated by replacing the three structures by their registered CAD model representation. The remaining points representing the seabed are replaced by a Poisson surface. As illustrated in fig. 7, the generated world representations (fig. 7-b-c) offer a continuous representation of the sampled scene (fig. 7-a) at a custom level of detail. While fig. 7-b offers an equivalent continuous representation of the initial reconstructed point cloud (500Mb), the representation exhibited in fig. 7-c only requires 5Mb.

V. DISCUSSION

A. Performance

We obtained our experimental results on a recent hardware configuration (Intel i7-4700MQ processor with 16GB RAM). Although the reconstruction of the full field in 3D can be obtained in a similar time to the full inspection time (1h30), the computing resource usage of the current implementation would make a vehicle integration for online mapping difficult. An adapted version of the algorithm aiming at the generation of a direct elevation map would however enable online processing.

The identification of the two first structures was obtained in 30sc at $\Delta_{NE} = 8$cm while the last and larger structure required up to 9mn. Similarly the model registration of each structure took from 1 to 15mn. These durations depend on

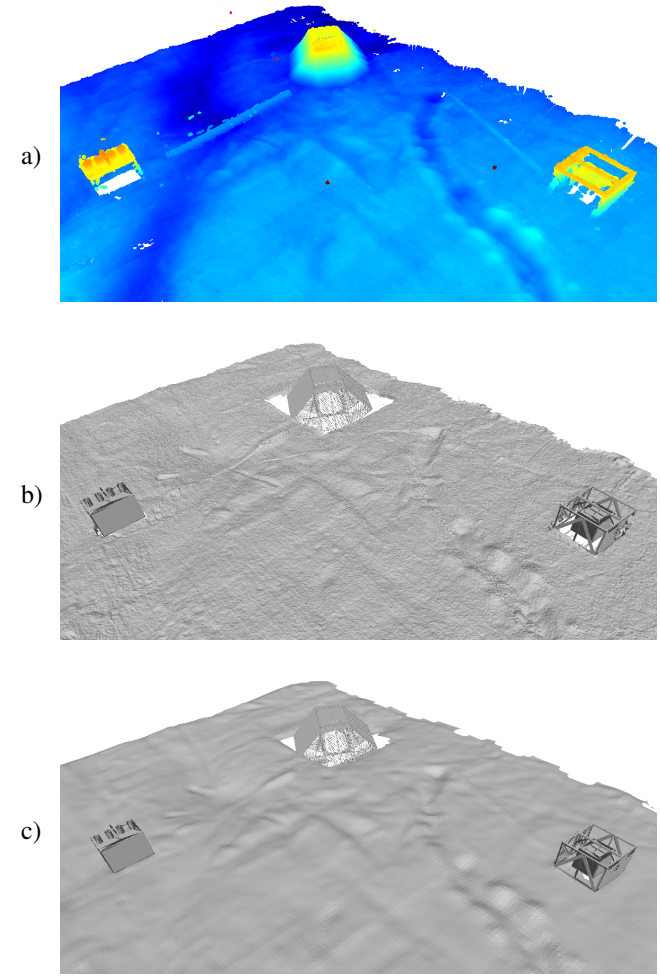


Fig. 7. Illustration of World model compression. The color-coded reconstructed point cloud (6.97M points) of the full field (a) can be represented by the set of registered CAD models and a surface representation of the seabed generated at various levels of details : b) 5.95M points and 11.9M faces, c) 33k points and 65k faces.

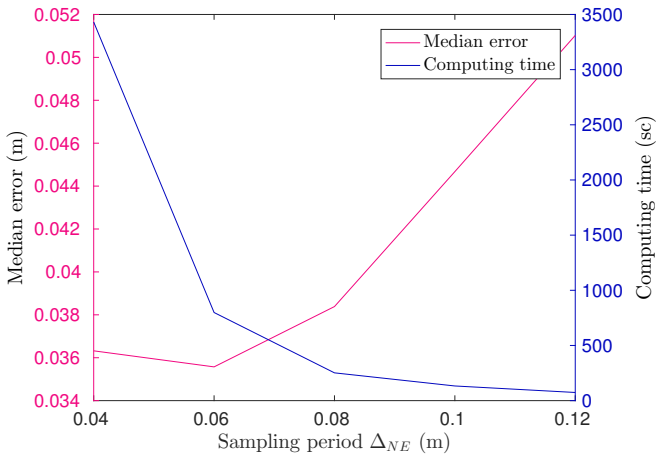


Fig. 8. Influence of the sampling period Δ_{NE} on error (magenta) and computing time (blue). The median distance between the model and the reconstruction after the structure recognition and the first registration step appears to be limited by the frame rate used during the survey (4cm sampling along track resolution). The computing time naturally increases with higher sampling resolutions.

the size of the structure, the sampling resolution and on the symmetry in the case of the first matching step. As illustrated in fig. 8, multi-resolution tests showed the robustness of the detection and matching performances at sampling resolutions ranging from 4 to 12cm. Similarly, the choice of the number of histogram bins did not appear determinant. Therefore and after further engineering, the choice of these parameters could be adapted to the specific situation based on the available computing resources, the size of the field, the size of the structures and the time constraints.

B. Limitations

Although two pipelines were present on the field of trials, the method described in this paper only enabled to identify and match accurately the biggest one of the two (25cm radius vs 10cm radius). This is due to the high eccentricity of the pipelines resulting in circular patches mainly filled by seabed points. Therefore the percentage of points representing the structure remains very low, leading to a noisy description of the object. In the case of a pipeline, a simple line detection algorithm applied to the elevation map of the field would be more adapted to the description of this specific geometry. The knowledge of the pipeline position provides connectivity information, adding semantic knowledge on the scene.

VI. CONCLUSIONS

This paper presented an underwater semantic mapping pipeline based on multibeam SONAR observations and a set of CAD models representing the structures of interest. A custom-size simplified 3D representation of the world is then generated for both online and post-mission use. This approach not only enables a semantic interpretation of the scene in operation time, paving the way to more complex operations such as manipulation, short-range or optimized structure inspection but also enables the generation of precise field maps addressing directly the common issue of inaccurate field information in the offshore industry.

Future work will focus on pipeline detection to allow for the observation of connectivity between the structures as well as the integration of video data in order to allow for direct visual inspection of the 3D space by an operator.

ACKNOWLEDGMENT

The authors would like to thank Subsea 7 for providing access to the AIV data. This project was funded under the Marie Curie ITN program Robocademy FP7-PEOPLE-2013-ITN-608096. Kartic Subr was supported by a Royal Society University Research Fellowship.

REFERENCES

[1] M. D. Aykin and S. Negahdaripour, "Three-dimensional target reconstruction from multiple 2-d forward-scan sonar views by space carving," *IEEE Journal of Oceanic Engineering*, 2016.

[2] C. Beall, B. J. Lawrence, V. Ila, and F. Dellaert, "3d reconstruction of underwater structures," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010, pp. 4418–4423.

[3] P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in *Robotics-DL tentative*. International Society for Optics and Photonics, 1992, pp. 586–606.

[4] A. Birk, K. Pathak, N. Vaskevicius, M. Pfingsthorn, J. Poppinga, and S. Schwertfeger, "Surface representations for 3d mapping," *KI-Künstliche Intelligenz*, vol. 24, no. 3, pp. 249–254, 2010.

[5] V. L. Chen, M. A. Batalin, W. J. Kaiser, and G. Sukhatme, "Towards spatial and semantic mapping in aquatic environments," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*. IEEE, 2008, pp. 629–636.

[6] M. F. Fallon, J. Folkesson, H. McClelland, and J. J. Leonard, "Relocating underwater features autonomously using sonar-based slam," *IEEE Journal of Oceanic Engineering*, vol. 38, no. 3, pp. 500–513, 2013.

[7] T. Guerneve and Y. Petillot, "Underwater 3d reconstruction using blueview imaging sonar," in *OCEANS 2015-Genova*. IEEE, 2015, pp. 1–7.

[8] T. A. Huang and M. Kaess, "Towards acoustic structure from motion for imaging sonar," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE, 2015, pp. 758–765.

[9] M. V. Jakuba, D. Steinberg, J. C. Kinsey, D. R. Yoerger, R. Camilli, O. Pizarro, and S. B. Williams, "Toward automatic classification of chemical sensor data from autonomous underwater vehicles," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*. IEEE, 2011, pp. 4722–4727.

[10] J. W. Kaeli, J. J. Leonard, and H. Singh, "Visual summaries for low-bandwidth semantic mapping with autonomous underwater vehicles," in *Autonomous Underwater Vehicles (AUV), 2014 IEEE/OES*. IEEE, 2014, pp. 1–7.

[11] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson Surface Reconstruction," in *Symposium on Geometry Processing*, A. Sheffer and K. Polthier, Eds. The Eurographics Association, 2006.

[12] A. Kenny, I. Cato, M. Desprez, G. Fader, R. Schüttenhelm, and J. Side, "An overview of seabed-mapping technologies in the context of marine habitat classification," *ICES Journal of Marine Science: Journal du Conseil*, vol. 60, no. 2, pp. 411–418, 2003.

[13] I. Kostavelis and A. Gasteratos, "Semantic mapping for mobile robotics tasks: A survey," *Robotics and Autonomous Systems*, vol. 66, pp. 86–103, 2015.

[14] G. Marani, S. K. Choi, and J. Yuh, "Underwater autonomous manipulation for intervention missions auvs," *Ocean Engineering*, vol. 36, no. 1, pp. 15–23, 2009.

[15] T. Masuda, K. Sakaue, and N. Yokoya, "Registration and integration of multiple range images for 3-d model construction," in *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, vol. 1. IEEE, 1996, pp. 879–883.

[16] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, "Hybrid genetic optimization and statistical model based approach for the classification of shadow shapes in sonar imagery," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 2, pp. 129–141, 2000.

[17] K. Mizuno and A. Asada, "Three dimensional mapping of aquatic plants at shallow lakes using 1.8 mhz high-resolution acoustic imaging sonar and image processing technology," in *Ultrasonics Symposium (IUS), 2014 IEEE International*. IEEE, 2014, pp. 1384–1387.

[18] A. Palomer, P. Ridao, and D. Ribas, "Multibeam 3d underwater slam with probabilistic registration," *Sensors*, vol. 16, no. 4, p. 560, 2016.

[19] D. Pangercic, B. Pitzer, M. Tenorth, and M. Beetz, "Semantic object maps for robotic housework-representation, acquisition and use," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE, 2012, pp. 4644–4651.

[20] M. Pfingsthorn, A. Birk, and N. Vaskevicius, "Semantic annotation of ground and vegetation types in 3d maps for autonomous underwater vehicle operation," in *OCEANS 2011*. IEEE, 2011, pp. 1–8.

[21] A. Pronobis and P. Jensfelt, "Large-scale semantic mapping and reasoning with heterogeneous modalities," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 3515–3522.

[22] D. Ribas, P. Ridao, J. D. Tardós, and J. Neira, "Underwater slam in man-made structured environments," *Journal of Field Robotics*, vol. 25, no. 11-12, pp. 898–921, 2008.

[23] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *International journal of computer vision*, vol. 40, no. 2, pp. 99–121, 2000.

[24] I. T. Ruiz, Y. Petillot, D. Lane, and J. Bell, "Tracking objects in underwater multibeam sonar images," 1999.

[25] M. VanMiddlesworth, M. Kaess, F. Hover, and J. J. Leonard, "Mapping 3d underwater environments with smoothed submaps," in *Field and Service Robotics*. Springer, 2015, pp. 17–30.